NOAA Technical Memorandum OAR-GSD-32

THE INITIAL FORMULATION OF A TECHNIQUE TO EMPLOY GRADIENT INFORMATION IN A SIMPLE VARIATIONAL MINIMIZATION SCHEME

Daniel Birkenheuer

Earth System Research Laboratory
Global Systems Division
Boulder, Colorado
January 2006

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Office of Oceanic and Atmospheric Research

NOAA Technical Memorandum OAR-GSD-32

THE INITIAL FORMULATION OF A TECHNIQUE TO EMPLOY GRADIENT INFORMATION IN A SIMPLE VARIATIONAL MINIMIZATION SCHEME

Daniel Birkenheuer
NOAA Earth System Research Laboratory, Boulder, CO

Earth System Research Laboratory
Boulder, Colorado
January 2006

NOTICE

For sale by the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22061

CONTENTS

The Initial Formulation of a Technique to Employ Gradient Information in a Simple Variational Minimization Scheme

Daniel Birkenheuer

ABSTRACT

In 1989, when the first version of the Local Analysis and Prediction System (LAPS) emerged from the drawing boards, it was equipped with a gradient analysis component to utilize satellite data structure. At that time, satellite data were not well calibrated and ancillary data were used to "anchor" the data field, while satellite gradients were then applied between these points to fill in structure. Since the launch of *GOES-8*, better blackbody calibration on the spacecraft and improved ground comparison (and moon-look) techniques were implemented to reduce, in theory, the overall error of the satellite data. The LAPS system then dropped its gradient approach for satellite data in favor of direct assimilation as FSL began incorporating 3DVAR techniques in its analyses, based on the assumption that observation errors were well defined and bias was negligible.

Then in 2002, a landmark water vapor experiment, the International H_2O Project (IHOP), allowed the Forecast Systems Laboratory (FSL) the opportunity to compare GOES retrieved integrated water vapor with the equivalent measurement made from a set of ground-based, satellite independent measurements from the radio signal delay data using Global Positioning System (GPS) data. The results revealed a significant moist bias in the satellite measurements at asynoptic times. It became apparent that the GOES retrievals were optimized for synoptic times and these corrections were not working well for the majority of the GOES products at intervening time periods.

There are two independent paths to address this problem. One is to correct the GOES product itself, and that is being pursued and is not the subject of this paper. The second approach, and the theme of this paper, is to modify the assimilation system to a form that only utilizes satellite gradients, much like the initial approach in LAPS years ago. To initiate this change, some preliminary work had to be undertaken to ascertain the weights to be applied to gradient terms in the revised analysis functional. That is the main topic of the experiment described in this document. To do this, an analytical function was used to describe "truth," and a gradient version of this function was used for "satellite gradient observations;" then a degraded version of the function was used for a "background field." Numerical experiments were undertaken to measure the effectiveness in the gradient information in minimizing the error by using variational analysis to improve the background field to fit truth.

Once the relative coefficient weights were determined in the foregoing analytic test, the new numerical analysis equations were incorporated into the existing LAPS software and compared to real data cases. One comparison (randomly selected) is presented in this memo. In all cases examined, the impact was significant; bias error was remarkably reduced and analyzed structural detail was vastly improved in moist areas. This report

includes Appendix B that documents the actual changes to the LAPS FORTRAN software that is now being used operationally.

1. Introduction

In 1989, when the first version of the Local Analysis and Prediction System (LAPS) initially came on line (McGinley et al. 1991), the moisture analysis essentially modified a model background using Horizontal Shape Matching (HSM) (Birkenheuer 1996) to integrate satellite structure into the moisture analysis. This approach was taken primarily in response to the fact that the satellite data was from the Visible Spin Scan Radiometer Atmospheric Sounder (VAS) instrument predating the three axis stabilized geostationary operational environmental satellite *(GOES)-8*, and had better on-board calibration capabilities. HSM was based on a minimization technique imposing weak constraints to merge satellite gradient structure with more accurate ground-based data. The VAS instrument was poorly calibrated compared to more modern weather satellites, and as such, continual bias issues had to be dealt with in order to make best use of the data. HSM was but one means of dealing with bias.

In the mid-1990s, the moisture module in the LAPS system transitioned to a variational minimization (1DVAR) operating at each individual gridpoint (or in the interests of speed, every second or third point using interpolation to achieve a finished product) to combine the diverse data sources that were added to the assimilation in the late 1990s and early 2000s. At about this same time, *GOES-8* data products became available, and it was assumed that these products would have minimal bias error due to the better onboard black-body calibration techniques available with the new satellite series. The variational system at that time dropped gradient assimilation in lieu of directly using derived product total precipitable water (TPW) values in concert with other new moisture data sources such as global positioning system (GPS), GOES direct radiance data through the use of a newer forward model (OPTRAN), and the inclusion of cloud information in the solution (Birkenheuer 1991).

This approach to moisture analysis was deemed most favorable until about 2004 following an extensive reanalysis of the moisture data from the International H_2O Project (IHOP)–2002 (Birkenheuer and Gutman 2005). It was observed that the published low bias figures from satellite product moisture used in the analysis system were valid for only 0000 UTC, and that asynoptic times as well as 1200 UTC GPS-GOES comparisons showed much greater GOES moist bias than had been perceived to exist. Since modern variational methods are highly dependent on both bias and variance/covariance error statistics, this recent finding has far-reaching effects.

There are two independent strategies to cope with a GOES bias problem; one could be directly addressed by product developers (i.e., make a better product), and in addition, a second was to modify the analysis system to ignore bias and focus on structure in a similar way to HSM. Both approaches were exclusive, so the Forecast Systems Laboratory, in cooperation with its Development Division, has established a web page that helps to compare the GOES product with both RAOB and GPS data in real-time.

This web page has already established beyond a doubt that the GOES product continues to contain unexplained bias on the order observed during IHOP 2002. It is being used by GOES product developers to study and address the bias problem. Figure 1 compares a daily difference of GOES and GPS data for 12 days. A periodicity is readily seen that reinforces the idea that the GOES bias is best at 0000 UTC exactly in line with IHOP results.

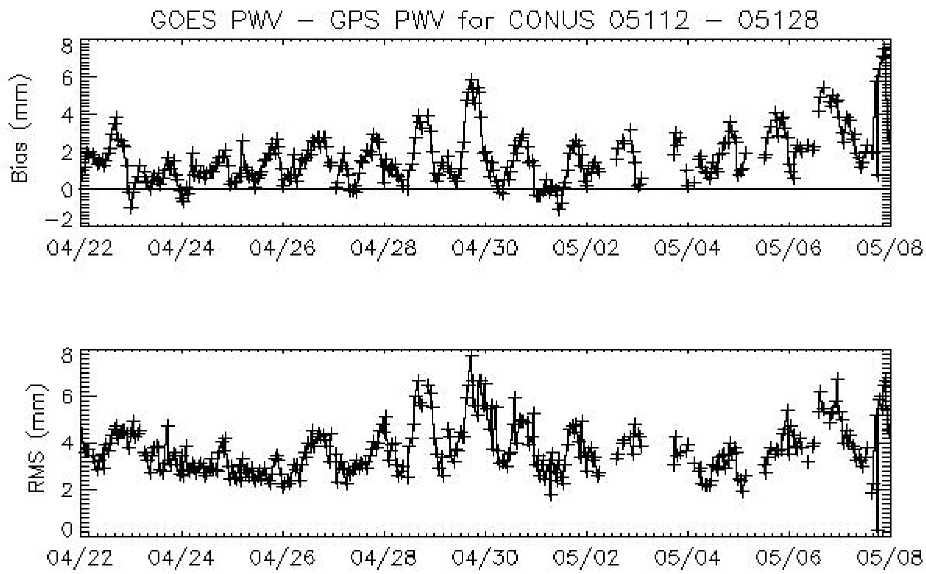

Fig. 1. Shows the GOES-GPS difference (mm "bias") top and the RMS (mm) below from late April into May 2005. The periodic nature of the derived moisture is evidenced in the top figure where "bias" values drop near zero at evenly spaced intervals. This was the identical behavior noted during IHOP–2002. The computation here, however, is derived from GOES 12 and all of the GPS sites in the eastern half of the continental US, a much larger region than IHOP, but for a shorter time period.

The second approach is the theme of this technical report. It documents just how the variational algorithm that has served the moisture product for numerous years has now been adjusted from a 1DVAR system to one that remains 1DVAR but contains data from surrounding gridpoints to assimilate horizontal gradient information that is now automatically incorporated in the minimization processing. By minimizing partial derivatives in orthogonal directions, the assimilation system both incorporates the gradient structure inherent in high-resolution satellite product data while becoming immune to bias problems.

2. Theory

The theory for this application goes back to the older HSM method that minimized the partial x and y derivatives of the observed field to the corresponding derivative of the background. By taking the partial derivatives, we essentially eliminate bias since a measurement bias is a constant and derivatives of constants are zero. The premise is that adding gradient data from satellite *image* products or single field of view sounding channels will be superior to supplying the measurement to the minimization, since the bias in the product data (especially the single field of view data) has been shown to be greater than anticipated at asynoptic times (Birkenheuer and Gutman 2005). Furthermore, by minimizing to gradients, especially when the gradient is at greater resolution than other measurements, one gains a solution that will contain maximum gradient information. This is where an instrument like satellite data really come through is that they may have flawed accuracy, but for a given scan, the relative measurements in the horizontal reflect the structure of the field.

Bringing this idea into a minimization scheme like 1D or 3DVAR, one retains the terms for absolute measurements such as GPS that are horizontally sparse while adding terms that minimize the partial derivatives of the solution to the partial derivatives of the satellite field. This produces a solution that will be as accurate as possible in theory to absolute data sources such as RAOBS (if they are timely), GPS, and any other quality point data along with the strength of the satellite relative gradient data structure.

In this application, the adaptation to the gradient minimization of satellite structure was fairly straight forward since the satellite product moisture term was simply transformed into its corresponding partial derivative form. The satellite gradient terms (terms 5 and 6) are denoted in the following equation that is the one proposed for the new analysis scheme.

$$
\begin{aligned}
J = {} & S_{SAT} \sum_{k=1}^{7} \frac{\mathrm{GT}(g_i)[\mathrm{R}(T, cq, o_3)_i - R_i^o]^2}{E_{SAT}^2} \quad + \quad \sum_{i=1}^{N} \frac{(1-c_i)^2}{E_{BACK}^2} \\
& + S_{GPS} \frac{\left(\sum_{i=1}^{N} c_i q_i - Q^{GPS} \right)^2}{E_{GPS}^2 L_{GPS}} \quad + \quad S_{sonde} \frac{\sum_{i=1}^{N} \left[\mathrm{RH}(T, p, cq)_i - RH_i^o \right]^2}{E_{sonde}^2 L_{sonde}} \\
& + S_d S_{GVAP} \sum_{j=1}^{3} \frac{\mathrm{G}(g)[\sum_{i=1}^{N} \frac{\Delta}{\Delta x} \mathrm{P}_{ji}(c_i q_i) - \frac{\Delta}{\Delta x} Q_j^{GVAP}]^2}{E_{xGVAP_j}^2 L_{GVAP}} \quad \text{(Term 5)} \\
& + S_d S_{GVAP} \sum_{j=1}^{3} \frac{\mathrm{G}(g)[\sum_{i=1}^{N} \frac{\Delta}{\Delta y} \mathrm{P}_{ji}(c_i q_i) - \frac{\Delta}{\Delta y} Q_j^{GVAP}]^2}{E_{yGVAP_j}^2 L_{GVAP}} \quad \text{(Term 6)} \\
& + \quad S_{CLD} \sum_{i=1}^{N} \frac{\mathrm{g}_i [c_i q_i - q_s(t_i)]^2}{E_{CLD}^2}
\end{aligned}
$$

The terms in the above equation are defined in Birkenheuer 2001 and also appear in Appendix A with the exception of the new partial derivative terms, S_d, (the switch for the derivative [0=off, 1=on]}, and the error terms subscripted with x and y are in regard to gradient error. Gradient error is a very little understood quantity at this point and will require future study and testing.

3. Testing the Formulation

There are various ways to go about testing the above formulation and drawing on past experience; it was decided the easiest approach was to isolate the terms of interest and perform a minimization analysis on an analytic function for which one could easily compute both "truth" and numeric derivatives, thus simulating real practice. This test system would help answer the critical question, "what should the relative weight for a gradient term be in comparison to the measurement terms?" Having now familiarity with how gradient terms would "behave" in a variation system, the following equation was used to simulate "truth" and hypothetical, background, GOES, and GPS data terms in this simplified functional for testing; a series of numerical tests could easily be conducted using the following model.

$$Truth\ \ data = T = T(x, y)$$

$$T(x, y) = 12.5\,[\sin(x\frac{\pi}{4}) + \sin(y\frac{\pi}{4})] + x^2 + y^2$$

For this example, the sine term was included to simulate some kind of variability in the field that was about 1% of the value mid-way in the interval studies (1 – 50). The testing was performed in two parts; first was to examine only the last two squared terms as the truth function followed by a more complex test using the full equation that included the sine term.

The satellite gradient data was taken directly from the truth data with the hypothetical (best case) initial approximation that the satellite data was a *perfect* gradient measurement.

$$\frac{\partial}{\partial x}T(x, y) = 12.5\frac{\pi}{4}\ \cos(x\frac{\pi}{4}) + 2x$$

$$\frac{\partial}{\partial y}T(x, y) = 12.5\frac{\pi}{4}\ \cos(y\frac{\pi}{4}) + 2y$$

Numerically for both the truth data and its partial derivatives, it was evaluated using:

$$x = i$$

$$y = j$$

The background field for the study was initially set as an inferior match to the truth function:

$$Background = B(x, y) = x^{1.8} + y^{1.8}$$

Thus, the background is not a perfect fit to the squared power law followed by the truth data nor does it reflect the sinusoidal characteristics of the added "structure" incorporated in the "truth" field.

The partial derivatives from the background field were computed, as they would normally be in a real case, by differencing:

$$\frac{\Delta B(x,y)}{\Delta x} = 0.5[B(x_{i+1}, y_j) - B(x_{i-1}, y_j)] \approx B_x^{'}(x,y)$$

$$\frac{\Delta B(x,y)}{\Delta y} = 0.5[B(x_i, y_{j+1}) - B(x_i, y_{j-1})] \approx B_y^{'}(x,y)$$

The above differences breakdown at the boundaries, and these were simply defined as zero in this simulation. Primed variables are first order derivatives, and the subscript designates the variable of the derivative (not the constant variables). Furthermore, the boundary regions were ignored in the study of error. In actual application, this same approach could be taken (i.e., not to include satellite gradient effects at the boundary) or a simple approach of replicating the gradient adjacent to the boundary could be applied.

The minimized functional for evaluation was then formulated by the relationship using the above terms; it should be noted that this variational solutions differs from the one used for operational moisture analysis. Here we are only interested in determining the relative coefficient weights c_1 and c_2 so they can be applied in the moisture minimization later:

$$\begin{aligned} J = {} & c_1[p(x,y)B(x,y) - B(x,y)]^2 + c_2[p(x,y)B_x^{'}(x,y) - T_x^{'}(x,y)]^2 \\ & + c_2[p(x,y)B_y^{'}(x,y) - T_y^{'}(x,y)]^2 \end{aligned}$$

Where again, the primed functions are partial derivatives, either numerically or analytically generated. The "*C*" terms are empirical weights that control their effect relative to each other. The variable $p(i,j)$ is computed at each gridpoint to minimize J. Thus, when complete, the analyzed field (A) is computed as a modified background modulated by the "p" function:

$$A(x,y) = p(x,y)B(x,y)$$

or in terms of i and j,

$$A(i,j) = p(i,j)B(i,j)$$

Since the numerical derivates could not be computed at the boundaries, they were not included in a measure of error or fit, which was defined 4 points in from the boundary:

$$Error = \sum_{i=5}^{46}\sum_{j=5}^{46}[T(i,j) - A(i,j)]^2$$

No attempt to normalize the result is performed; instead the lowest number for the above computed error is taken to indicate the best fit to the data. Different sums are used to ratio results and gauge effectiveness.

What follows are some of the results of these tests.

3.1 Experiment 1 (ignore sine terms):

The first experiment (and all subsequent tests) kept the gradient term (C2) at a value of 1.0 while varying the weight of the background term (C1). The error results were summed within five gridpoints inside the boundary; the total domain was [50,50]. Relative improvement was computed from the background-only term that was run one time where C1 equaled unity, and all other coefficients were assigned zero so there were no other effects on the solution (the first line in each of the tables below). The truth term for the first simple experiment was, $T(x, y) = x^2 + y^2$, with the associated partial derivatives $2x$ and $2y$, thereby simplifying the test by eliminating the sine terms.

C1		Error (no adjustment)	Error (w/ grad)	Percent
1	(C2=0)	39288.3		100.00%
1E-06	(C2=1)		8338.26	21.22%
0.00001	"		8016.21	20.40%
0.0001	"		5141.12	13.09%
0.001	"		12423.1	31.62%
0.01	"		33284.2	84.72%
0.1	"		38586.2	98.21%
1	"		39216.8	99.82%

3.2 Experiment 2: (addition of GPS data)

The next experiment added GPS data from the truth function at every 10^{th} gridpoint with exponential weighting drop-off to 0.5 at about 2 gridpoints from the central data point representing the measurement. The functional was modified slightly to add a GPS term that differenced the absolute modified background and the weighted truth data (GPS) at all points. This third term received a coefficient C3. Here C2 was again held to 1.0 (the minimum above) and C3 and C1 were varied for the GPS term.

C3/1		Error (no adjust)	Error (w/ grad w/GPS)	Percent
(C3=0)	(C2=0,C1=1)	39288.3		100.00%
1E-06	(C2=1)		5114.98	13.36%
0.00001	"		4890.07	12.77%

0.0001	"	3917.42	10.23%
0.001	"	19756.8	51.60%
0.01	"	45173.5	117.98%
0.1	"	51977.8	135.75%
1	"	52799.1	137.90%

It is interesting that the addition of the synthetic GPS data with an equal weight to the background producing the gradient minimum resulted in the greatest improvement. Weighting the GPS as 1.0, equal to the background data, resulted in an error greater than just using an unadjusted background. This indicates that there would be more to be gained by tuning the system to the GPS data if it were to be used alone. In this case, it either had the best effect with the same coefficient value of the gradient term, or its biggest effect at some other weight was simply overwhelmed by the possible greater impact of the gradient assimilation. In actual application, it would likely be dependent on the functions (background and the GPS true distribution), which would likely change per each analysis, but we can use the coefficient values from experiments like this one to get an idea of a reasonable assignment for the coefficient values. We see that the lowest error with gradient term was about 13% of the pure background, a huge improvement in error over using the background field alone with no input data. That number only gains another 3% with the GPS data added to the field with the weight assigned here (no experiments were attempted varying the spatial weights of the GPS data itself). However, we also see that with a large weight on the GPS (0.01 or greater), the error rises above using the gradient information alone. This indicates that more can be gained in using gradient information with the raw GPS data.

3.3 Experiment 3: (include sine structure and exclude GPS)

The next test included more structure. It was unexpected that addition of the gradient term to the fairly smooth analysis field would have such a great impact. A more realistic test would be to include higher frequency structure between GPS point values, and the objective here would be to ascertain whether the gradient approach would further contribute to error reduction with more complex structure.

The set of measurement simulations used the added sinusoidal "variability" in the truth field shown at the beginning of this report. This amplitude was taken to be about 1% mid-way up the squared function curve so it is really about 100% variability at the low end and very small variability at the high end (large x, y values).

$$Truth\ \ data = T = T(x, y)$$

$$T(x, y) = 12.5\,[\sin(x\frac{\pi}{4}) + \sin(y\frac{\pi}{4})] + x^2 + y^2$$

and the known gradients are:

$$\frac{\partial}{\partial x}T(x,y) = 12.5\frac{\pi}{4}\cos(x\frac{\pi}{4}) + 2x = T_x^{'}(x,y)$$
$$\frac{\partial}{\partial y}T(x,y) = 12.5\frac{\pi}{4}\cos(y\frac{\pi}{4}) + 2y = T_y^{'}(x,y)$$

The quarter π ratio was selected such that a complete wave could be represented by 4 points or 40km in this simulation (assuming 10km grid spacing). This is less than the GPS spacing, and is nearly what one would get from the current GOES with a ~10 km sounder resolution.

C1		Background	Error using grad	Percent
1	(C2=0)	39291.85		100.00%
0.0000001	(C2=1)		9754.75	24.83%
0.000001	"		9501.94	24.18%
0.0001	"		7465.42	19.00%
0.001	"		13613.00	34.65%
0.01	"		33404.70	85.02%
0.1	"		38600.30	98.24%
1	"		39221.32	99.82%

The above results were similar to the field without the sinusoidal component. The gradient term with a weight of 1.0 and a background term of 0.0001 produced the best error reduction of about 80% (19% of the error).

3.4 Experiment 4: (sine structure plus GPS)

An additional run similar to the last test was made where GPS data was added to the functional. The tabulated results are shown next.

GPS data were again used from the truth function at every 10^{th} gridpoint with exponential weighting drop-off to 0.5 at about 2 gridpoints from the central data point representing the measurement. C1 and C3 were varied while holding C2 (gradient weight) at unity.

C3/1		Background	Error using grad	Percent
0	(C2=0,C1=1)	39291.85		100.00%
0.000001	(C2=1)		7442.960	18.94%
0.00001	"		7299.795	18.58%
0.0001	"		6716.925	17.09%
0.001	"		20388.37	51.89%
0.01	"		45243.84	115.15%
0.1	"		51984.34	132.30%
1	"		52799.78	134.38%

Using the GPS data in combination with the gradient data improved the error reduction by about another 2%, similar to what was seen in the less complex field, but overall the less complex field has less error. Figure 2 illustrates the drop in analysis error as the weight applied to the gradient term in the functional is increased relative to the non-gradient term.

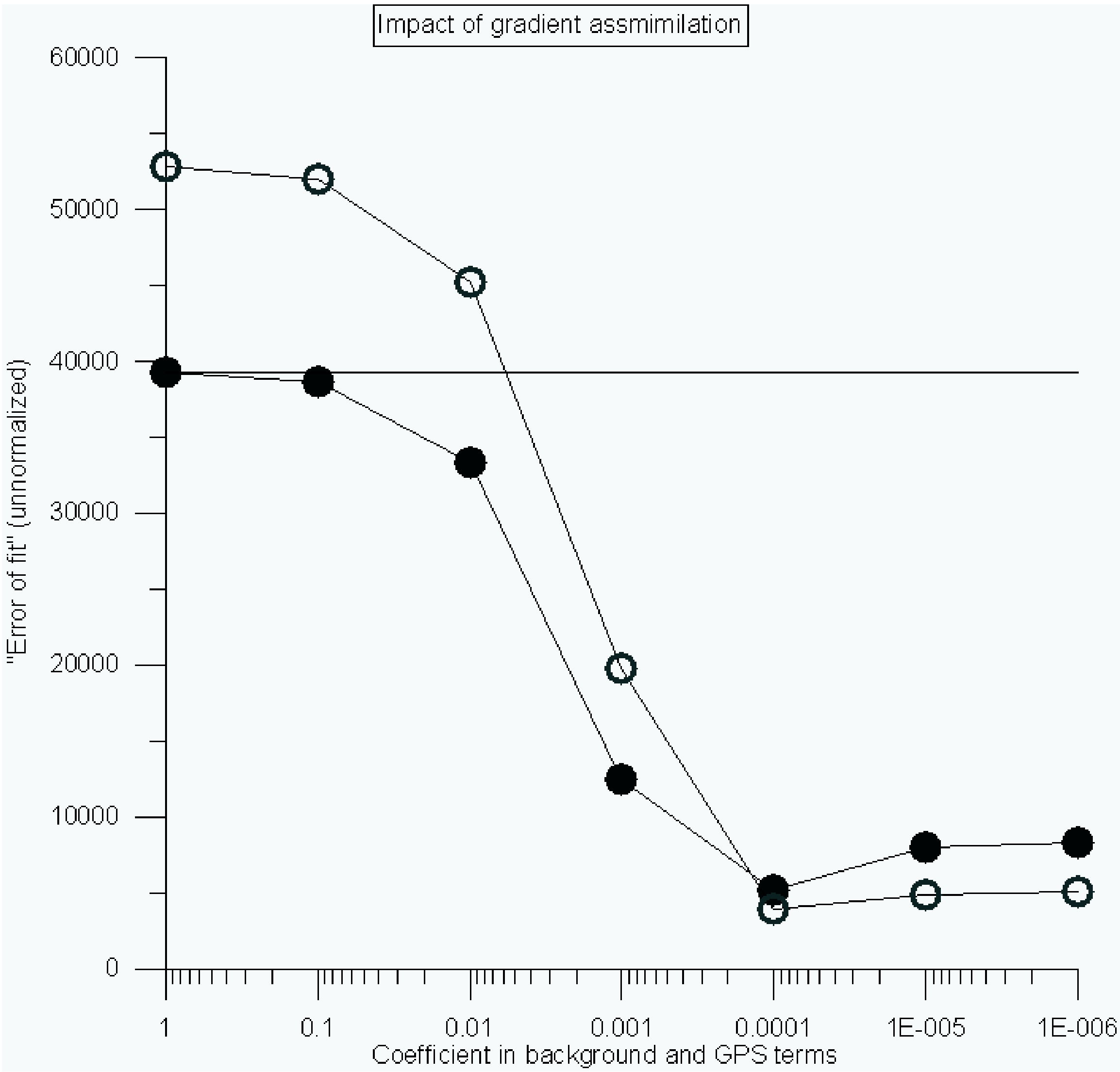

Fig. 2. Impact of simulated gradient data (solid circles); open circles also contain untuned GPS data results from experiment 2 (i.e., c_3 was unity in the related minimization equitation [not shown]).

It should be noted in all of these experiments that the error terms for the background, gradient, and GPS source data have been ignored. Simply stated, the problem was scaled down to its simplest form and again, is only useful to examine how various terms interact in the functional, whether the gradient approach has potential, and whether this technique is indeed a viable one for real-time application, since one of the actions that followed this initial testing was to code the system into the LAPS system and see just how well it performed both an analysis, and also in a computational sense to ascertain whether this approach would run in a reasonable amount of time in a real-time setting. It would be a realistic assumption that improved error information would have little if any degradation on the compute times involved.

3.5 Summary of analytic tests

In summary, the numeric analytical tests demonstrate that the gradient approach in a minimization configuration is tractable and requires little modification to existing code (an easy modification of the functional plus some preliminary computations of gradient fields are all that is required for implementation), solves the bias problem, and adds little to computational overhead with substantial reduction in error. It compliments the addition of GPS data in that non-gradient PW data is seen to not detract from the solution, but add to it. Any system, however, is not perfect and this, of course, is not without exception. Several problems for gradient application will be missing data, bad pixels, and clouds (as well as other data void areas or product artifacts that lead to false edges or unrepresentative gradients). Furthermore, partly cloudy regions may or may not have significant impact this remains a subject for additional study.

3.6 Clouds

Clouds will always obscure full-column infrared thermal and moisture measurements. For the application of satellite gradients in this application to total moisture, cloud problems don't go away; they simply morph into a different kind of problem. In the normal retrieval sense, clouds interfere with clear radiance and must be dealt with either by compensation or ignoring "contaminated" pixels; the tendency of clouds generally is to "cool" the radiance measurement over the true radiance representative of the ambient air temperature. In gradient work, the problem arises in that a partly cloudy pixel may instill a false gradient between a cloudy region and a clear region. In a retrieval, this would have the effect of causing a cold bias in the retrieved thermal profile; in the gradient application, it would possibly incorrectly place or modify the magnitude of thermal gradients. This may or may not present us with a big problem. If the partly cloudy features are only limited to cloud edges, it is probably not a big problem; in fact, it might lessen gradient "shock" that would transition from one region to another, but it could be further avoided by ignoring gradients near cloudy regions by placing a buffer-zone around clouds that would ignore gradient information. However, in an expansive area of partial cloudiness that is sub-pixel scale, what could happen is that perceived gradients in radiance will be more from fluctuation in the degree of cloudiness and not true gradients in moisture or thermal properties that one is after in the analysis. Therefore, in the first algorithm attempts, it might be best to error on the clear side and do some testing for partial cloudiness to avoid the latter type of pseudo gradients. Some day, data above cloud top would be used in the analysis, but for now we want to see the effect of clouds on the total field, thus clouds will represent data voids. These can be handled in the functional by toggling the gradient terms to zero in the area of gradient data voids. We can still allow total GPS data to exist in these areas. This would be an interesting experiment for a few issues.

One will be the effect of data void edges to the analysis. Edges bordering on zero data regions could be very high and if not treated correctly, would add a flawed signal to the analysis that could really cause uncontrollable error.

Another will be the effect that GPS data has in the center of the missing data areas. One might anticipate that the effects of GPS here should be greater than we have for the cloud-free tests above, since the gradient improvement will be reduced.

For the cloud test, a 50% cloud cover was approximated. This was done by using one large cloud in the middle of the domain so that it spans both the low and high signal parts of the field.

The findings were not surprising. This approach amounted to a simple fractional removal of the gradient term in the variational equation. Thus, in a linear manner, as more and more of the gradient is negated by clouds and thus not included in the solution, the solution error approaches the error by using only GPS data by the same fractional rate.

The real problem will occur at a cloud interface when the gradient computation might include data in the cloud region, say at point i+2, where there exists the possibility that i should not be used in the gradient computation. Therefore, for safety reasons, it seems prudent that in the application of the technique, there should be a region around the cloud-defined boundary that is "off-limits" for gradient computation and utilization. Therefore, if the cloud extends to j, the mask used to toggle off the gradient term will be extended to both j+2 (on the "high-j side of the cloud") and j-2 (on the low-j side of the cloud) for "safety" reasons.

Some computations were made for a simulated large cloud in the center of the above domain. The 6 716 "low value" in the last simulation (max error was 39 291) with an imposed cloud had an error value of 16 683, or 42% (note that the 6 716 had a 17% error, so the error more than doubled after adding cloud.) Another test was performed where the dimensions were extended as prescribed above as a buffer around the region the "error value" rose to 18 283 or 46.5%. So the additional buffer zone contributed to about 4.5% more error in the result. The question might be to see if this 4.5% could be reduced by other measures to avoid extreme gradient errors, but for a first step, this appears to be an acceptable price to pay for potential dire consequences of using false gradient information.

The above approach assumes that all clouds are well known and diagnosed properly with respect to coordinates i and j. However, this assumption may well break down in the real world, so some additional or alternate means to safeguard against spurious gradient data needs development.

4. Application of Gradient Method to Local Analysis System

The prescribed changes using a gradient approach to the LAPS variational moisture system were incorporated in March 2005. The following observations were encountered during this process, some of which were not envisioned in the initial plan.

1. The gradient field was a generation of surrounding gridpoints. At times, the surrounding values contained the LAPS missing data values (huge number 1e+37) that would cause a numeric fault in computation. This was actually good in that it pointed out the problem. The coding measures to circumvent this problem involved checking the horizontal field for these missing data flags and skipping the gradient step in the functional when they occurred.

2. The computation of the background layer gradient ran into problems when surrounding points were under ground and contained missing data flags. Then when the layer gradient was computed (and averaged), it might contain influence from a nearby missing data flag. This was coded around by making sure the gradient was under 1000. Generally, gradients were on the order of 1.0e-5, so this is well above that value and a good value to protect from missing data flag contamination. This problem would disappear if a terrain-following coordinate system was used, but LAPS traditionally uses an isobaric vertical coordinate.

3. The first set of comparison products were made between the LAPS parallel and operational runs. The parallel runs contained the gradient method, and the operational runs did not. Figures 3 and 4 contrast the results.

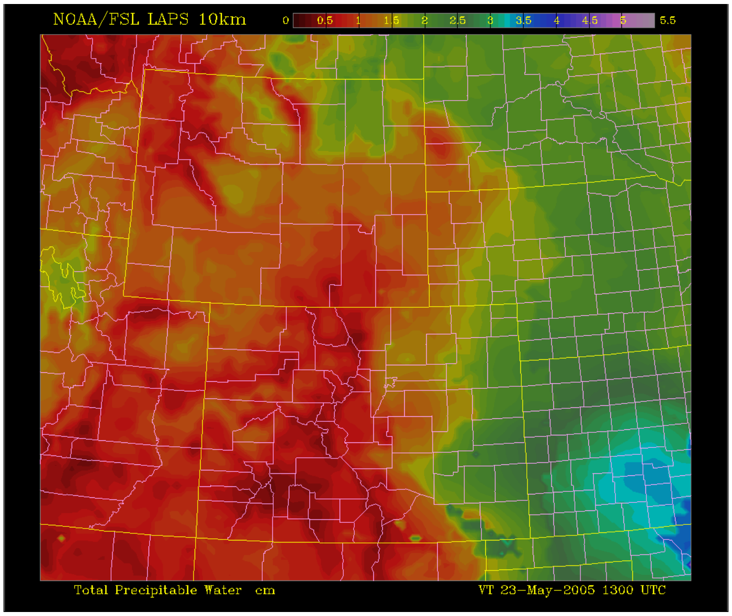

Fig. 3. Conventional analysis. Satellite data product used – direct use of satellite data potentially contaminated with moist bias for 23 May 2005 1300 UTC over Colorado and its surrounding states. Note that prior studies have demonstrated the maximum bias in the GOES product data occurs at 1800 UTC. This 1300 UTC analysis is actually in a timeframe of a secondary bias minima. However, moist bias is observed to be reduced in a number of circled areas, and increased structure is observed in "boxed regions" in the following figure.

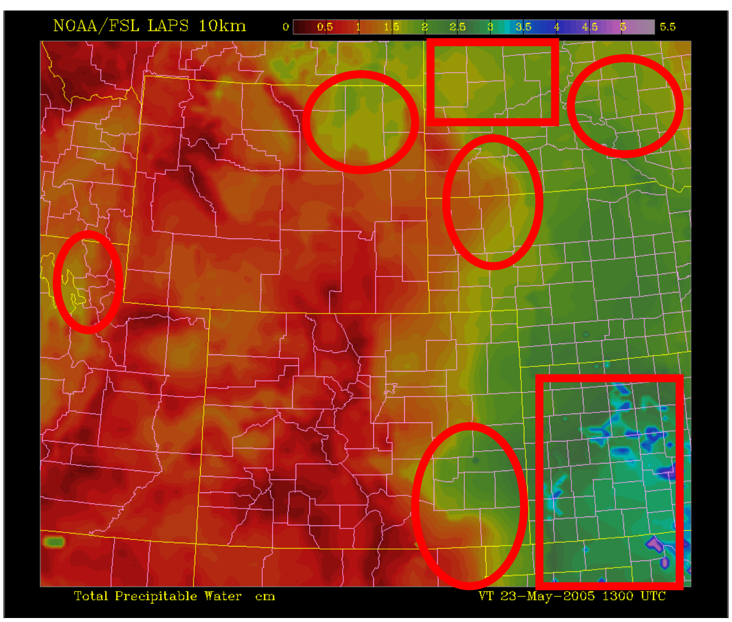

Fig. 4. Same as in Fig. 2 except using the new analysis relying on gradient information from satellite data and is disregarding satellite bias if it exists. The lower moist bias is evident in circled areas and better structural detail is evident in boxed regions.

Examining the two figures, it is evident that in the moist areas on the Kansas plains, there is more detail in the water vapor analysis with the gradient method. There also appears to be a good example of higher water vapor amounts in eastern Colorado in the conventional image (Fig. 3) indicative of a high moist bias coming from the satellite data, whereas Fig. 4 lacks this higher level and appears to render a more realistic field. A spectral analysis of both analyses (Fig. 5) reveals that the new technique not only eliminates satellite bias as mathematically mandated but provides more detail from the satellite data source.

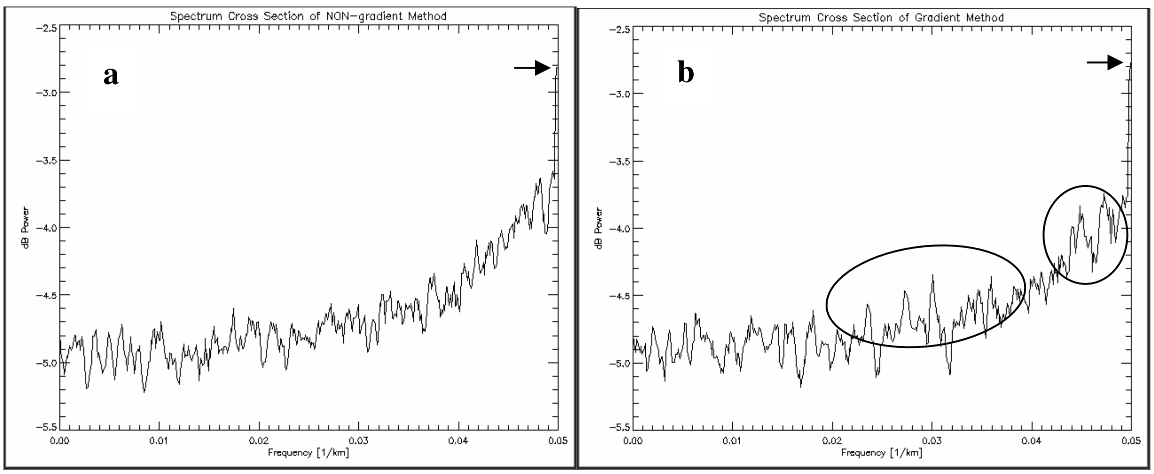

Fig. 5. Spectral plots showing respective power of image as a function of Fig. 3. (5a) and Fig. 4. (5b). The y-axis plots dB power, the x-axis plots frequency in Km^{-1}. It is evident due to the greater amplitudes at higher frequencies along with the greater variability in frequency amplitude in circled regions in 5b that Fig. 4 contains more structure. Also, in 5b, the highest frequency peak is slightly greater than in 5a, designated by arrows.

Also it should be noted that the dry areas between both analyses (high terrain areas) show the least difference, as one would expect since the GOES bias problems decrease with lower moisture amounts. This is only one case and should not be taken as "proof" that this technique works, however, it does demonstrate the attributes that the technique is designed to render.

5. Results and Conclusions

A numerical technique to include gradient information from satellite fields was successfully devised, first analytically and then cast into digital form for computer application. The analytical solution suggested the weights to apply to the gradient term in the functional used operationally in LAPS. The system was then run on real data comparing runs that both applied and did not apply the new technique, all other elements being equal.

Subjectively, it appears that the new methodology works both in reducing the moist bias inherent in the satellite data at this time and improving the structure in the resulting analyzed field. Figures 3 and 4 are highly representative of the routine effects that are observed when the new algorithm operates. The new method is now being run quasi-operationally in tests directly comparing the two methods. As of this writing, only positive results have appeared from the new technique as it is being readied for “operational” use in the “routine” LAPS analysis. The main purpose for this Technical Memo is to document this work and the changes to the LAPS software.

Appendix B contrasts the code changes that were needed in order to bring about this change in the LAPS system. Not shown in the Appendix are the filling of the partial derivative arrays; these are prepared and filled prior to calling the functional. In the analytic example above, the background gradients were computed numerically while the “satellite” gradients were perfect analytic derivatives. In practice, both the satellite and background partial derivatives are computed numerically.

All of the changes to the functional are not yet complete as this remains a work in progress; which is especially true for the partial derivative error terms for which there is no accepted value at the current time. It might be possible to discover the error terms by analytic methods similar to the tests conducted in this report. One could impose random error in a synthetic field and then from this, compute the partial derivative error estimates from this kind of setup.

6. References

McGinley, J.A., S.C. Albers, and P.A. Stamus, 1991: Validation of a composite convective index as defined by a real-time local analysis system. *Wea. Forecasting* **6**, 337-356.

Birkenheuer, D., 1996: Applying Satellite Gradient Moisture Information to Local-Scale Water Vapor Analysis using Variational Methods. *J. Appl. Meteor.* , **35**, 24-35.

Birkenheuer, D., and S. Gutman, 2005: A comparison of the GOES moisture-derived product and GPS-IPW during IHOP. *J. Atmos. Oceanic Tech.,* **22**, pp. 1840 – 1847.

Birkenheuer, D., 2001: Utilizing variational methods to incorporate a variety of satellite data in the LAPS moisture analysis, *11th Conference on Satellite Meteorology and Oceanography,* Amer. Meteor. Soc., Madison, WI, 273-276.

Birkenheuer, D.,1991: An algorithm for operational water vapor analyses integrating GOES and dual-channel microwave radiometer data on the local scale. *J. Appl. Meteor*. **30**, 834-843.

7. Appendix A: Term Definitions in Equation 1 from Birkenheuer 1991

$$J = S_{SAT}\sum_{k=1}^{7}\frac{GT(g_i)[R(t,cq,o_3)_i - R_i^o]^2}{E_{SAT}^2 L_{SAT}} + \sum_{i=1}^{N}\frac{(1-c_i)^2}{E_{BACK}^2}$$

$$+S_{GPS}\frac{(\sum_{i=1}^{N} c_i q_i - Q^{GPS})^2}{E_{GPS}^2 L_{GPS}} + S_{sonde}\frac{\sum_{i=1}^{N}\left[\mathrm{RH}\left(T,p,cq\right)_i - RH_i^o\right]^2}{E_{sonde}^2 L_{sonde}}$$

$$+S_{GVAP}\sum_{j=1}^{3}\frac{G(g)(\sum_{i=1}^{N} P_{ji}(c_i q_i) - Q_j^{GVAP})^2}{E_{GVAP_j}^2 L_{GVAP}} + S_{CLD}\sum_{i=1}^{N}\frac{g_i[c_i q_i - q_s(t_i)]^2}{E_{CLD}^2 L_{CLD}}$$

Each term in the above functional is modified by the variable *S*, which is a switch (with the exception of the background term which is always on). Thereby, the terms can be used or ignored depending on whether or not data are available or if clouds are present. Furthermore, a user can easily add terms for new datasets by simply creating a new term. An example of this is the RH term which was added since the initial publication of Birkenheuer 1991. The variables are as follows:

- C_i the coefficient vector applied to q to adjust the moisture field. Ideally this would have the same dimensions as q has levels, but may be reduced depending on computer horsepower. Adjustment of this parameter is in essence the variational fit to the solution, i.e., $c_i q$ becomes the adjusted q field. The adjustment coefficient is a scalar with a lower limit of 0 (never negative). A value of 1 indicates no change to the background. Because of this, the system will only work with a quantity such as temperature or humidity that uses absolute units. For example, using this approach to analyze temperature in degrees F will fail.
- q the specific humidity profile at one LAPS grid point
- R the forward-modeled radiance or radiance observation with the superscript o.
- RH relative humidity, measured or compute, either from sonde or derived satellite retrieval in some situations.
- i index for the LAPS vertical (vector dimension of q), with a current maximum of 40 (accommodating the climatological stratospheric layers needed for the forward radiance model).
- k the index indicating the satellite sounder or imager channel used.
- Q^{GPS} the total precipitable water measurement from GPS.
- E the error function (squared quantity) that describes the observation or background error, subscripted by observation type.

- *L* spatial weighting term subscripted by observation type. This weights the smoothed (preanalyzed) field value by its proximity to the observation and reflects the horizontal influences of the measurement. Each data source has an associated gridded field of spatial-weighting terms characterizing its proximity to the observation and its spatial representation.
- *P* the function to convert from pressure to sigma coordinates
- Q^{GVAP} the GOES vapor total precipitable water layer data. The layers are defined in sigma coordinates and vary grid point to grid point.
- *j* the index of the GVAP layer, with a current maximum of 3 (1 is lowest, 3 is highest).
- *Cld* cloud function designating cloudy regions in the vertical, with dimensions of *q*.
- *J* the functional to be minimized.
- *t* is the temperature profile (LAPS) at the same location as *q*.
- *S* logical switch for the observation type to be present or not. Each term in the functional can be easily included or excluded depending on the presence of the data source. Also new data sources can be added by including new terms.
- $q_s(t)$ saturated *q* as a function of temperature.
- *g* cloud fraction indicator as a function of level.
- *G* a function of *g* such that it indicates cloud in the column. For radiance measurements, this has the advantage of disabling IR terms including GVAP. Finally, the GPS term would be unaffected by clouds in principle since the data source can deliver data in cloudy areas. However, the analysis needs to probably give more credence to the cloud field, since it is vital the cloud field complements the moisture field. *G* can be a linear function of cloud such that it might serve to help define partly cloudy regions by allowing a smooth gradient from total through partly cloudy to clear air.

GT is a similar function to *G,* but it may be nonlinear and can match the satellite radiometer's field of view.

8. Appendix B: Software Changes for the Local Analysis and Prediction System Moisture Analysis Segment

The above gradient technique was coded into the appropriate functional of the Local Analysis and Prediction Systems (LAPS) moisture variational software. The routine that handles the functional (minimized equation, FUNC_O.F) was modified with the following modifications inserted to replace the direct use of the GOES information.

Older code with no gradient usage:

```
c     GVAP SECTION -- UNITS mm

      if (cost_gvap_istatus ==  1) then

c     test for weight of measurement
         if(cost_weight == cost_mdf) then !skip this step
            continue            ! skip this iteration
         else                   ! process gvap
            if (first_gvap) then
               first_gvap = .false.
               write(6,*) 'TEMP GVAP accepted'
            endif

c     integrate q for gvap layers
c     determine sigma level pressure analogs

            if (cost_gvap_p == cost_mdf) then
               cost_gvap_p  = cost_ps ! substitute for missing gvap press
            endif

            call sigma_to_p (0.1, cost_gvap_p, 0.9, p1)
            call sigma_to_p (0.1, cost_gvap_p, 0.7, p2)
            call sigma_to_p (0.1, cost_gvap_p, 0.3, p3)

            call int_layerpw(x,cost_data,cost_kstart,
     1           cost_qs,cost_gvap_p,cost_p1d,p1,p2,p3,lpw1,lpw2,lpw3,
     1           cost_kk,cost_mdf)

            if (p1 <= 300.0) then
               write(6,*)'TEMM ', x, p1,p2,p3,lpw1,lpw2,lpw3,
     1              cost_w1,cost_w2,cost_w3
            endif

            if (lpw2 == cost_mdf) then
               i = i
            endif

            max_func_gvap1 = 0.0
            max_func_gvap2 = 0.0
            max_func_gvap3 = 0.0

            if (lpw1 /= cost_mdf) then
               max_func_gvap1 =
     1              (lpw1-cost_w1)**2*cost_weight
            endif
            if (lpw2 /= cost_mdf) then
               max_func_gvap2 =
     1              (lpw2-cost_w2)**2*cost_weight
            endif
            if (lpw3 /= cost_mdf) then
               max_func_gvap3 =
     1              (lpw3-cost_w3)**2*cost_weight
            endif
```

```
c     lpw and cost are in mm tpw.  layer 1 is weighted very low since
c     it seems to always disagree with model computations by a significant
c     amount.  this are under research

c            max_func_gvap1 = 0.0  ! give layer 1 no influence

            max_func_gvap = (max_func_gvap1+max_func_gvap2
     1           +max_func_gvap3)

c     max_func_gvap is in mm (above) now divide by mm error to make
c     dimensionless
            max_func_gvap = max_func_gvap / ((3.27)**2) ! from SFOV worst case
            func = func + max_func_gvap
            if (max_func_gvap .ne. 0.0) then
               min_func_gvap = min(min_func_gvap,max_func_gvap)
            endif



         endif                  !weight function test
      endif                     !data present test
C     END GVAP SECTION
```

New Gradient Section:

```
c     GVAP SECTION -- UNITS mm

      if (cost_gvap_istatus ==  1) then

c     test for weight of measurement
         if(cost_weight == cost_mdf) then !skip this step
            continue            ! skip this iteration
         else                   ! process gvap
            if (first_gvap) then
               first_gvap = .false.
               write(6,*) 'TEMP GVAP accepted'
            endif

c     integrate q for gvap layers
c     determine sigma level pressure analogs

            if (cost_gvap_p == cost_mdf) then
               cost_gvap_p  = cost_ps ! substitute for missing gvap press
            endif

            call sigma_to_p (0.1, cost_gvap_p, 0.9, p1)
            call sigma_to_p (0.1, cost_gvap_p, 0.7, p2)
            call sigma_to_p (0.1, cost_gvap_p, 0.3, p3)

c     now broken into gradiants in x and y
c     first the x grad

            call int_layerpw(x,cost_grad_x,cost_kstart,
     1           cost_qs,cost_gvap_p,cost_p1d,p1,p2,p3,lpw1,lpw2,lpw3,
     1           cost_kk,cost_mdf)

            if (p1 <= 300.0) then
               write(6,*)'TEMM ', x, p1,p2,p3,lpw1,lpw2,lpw3,
     1              cost_w1_x,cost_w2_x,cost_w3_x
            endif

            if (lpw2 == cost_mdf) then
               i = i
            endif

            max_func_gvap1 = 0.0
            max_func_gvap2 = 0.0
            max_func_gvap3 = 0.0
```

```
            if (abs(lpw1) .lt.  1000.) then
               max_func_gvap1 =
     1              (lpw1-cost_w1_x)**2*cost_weight
            endif
            if (abs(lpw2) .lt.  1000.) then
               max_func_gvap2 =
     1              (lpw2-cost_w2_x)**2*cost_weight
            endif
            if (abs(lpw3) .lt. 1000.) then
               max_func_gvap3 =
     1              (lpw3-cost_w3_x)**2*cost_weight
            endif

            max_func_gvap = (max_func_gvap1+max_func_gvap2
     1           +max_func_gvap3)

c      now repeat for the y gradient

            call int_layerpw(x,cost_grad_y,cost_kstart,
     1           cost_qs,cost_gvap_p,cost_p1d,p1,p2,p3,lpw1,lpw2,lpw3,
     1           cost_kk,cost_mdf)

            if (p1 <= 300.0) then
               write(6,*)'TEMM ', x, p1,p2,p3,lpw1,lpw2,lpw3,
     1              cost_w1_y,cost_w2_y,cost_w3_y
            endif

            if (lpw2 == cost_mdf) then
               i = i
            endif

            if (abs(lpw1) .lt.1000.) then
               max_func_gvap1 =
     1              (lpw1-cost_w1_y)**2*cost_weight
            endif
            if (abs(lpw2) .lt.1000.) then
               max_func_gvap2 =
     1              (lpw2-cost_w2_y)**2*cost_weight
            endif
            if (abs(lpw3) .lt.1000.) then
               max_func_gvap3 =
     1              (lpw3-cost_w3_y)**2*cost_weight
            endif

            max_func_gvap = (max_func_gvap1+max_func_gvap2
     1           +max_func_gvap3) + max_func_gvap
            max_func_gvap = max_func_gvap * 10000. ! analytic importance

c      max_func_gvap is in mm (above) now divide by mm error to make
c      dimensionless
            max_func_gvap = max_func_gvap / ((3.27)**2) ! from SFOV worst case
            func = func + max_func_gvap
            if (max_func_gvap .ne. 0.0) then
               min_func_gvap = min(min_func_gvap,max_func_gvap)
            endif


         endif                  !weight function test
      endif                     !data present test
C     END GVAP SECTION
```

Note that the newer code has a repeated section to handle the partial derivatives in both *x* and *y* independently. The newer code also includes a factor of 10 000 to compensate for the *C1* adjustment in the analytic testing of 0.0001 that was found to be optimal (since in this functional, the coefficient term on the background remains unity, and other terms *[not shown]* have remained unchanged). The remaining adjustments that were found to be useful from the bias application were left in place such as the mm error adjustment.